Invited paper for Anniversary session at ECAI 2012, Montpellier August 2012
To be Published in a special issue of *AI Communications*
(This version, available at the Birmingham CogAff web site, is liable to be updated.)

Biological, computational and robotic connections with Kant's theory of mathematical knowledge

Aaron Sloman
School of Computer Science, University of Birmingham, UK
http://www.cs.bham.ac.uk/~axs

October 23, 2013

Abstract

Abstract: In my research I meander through various disciplines, using fragments of AI that I regard as relevant, willing to learn from anyone whose ideas contribute. This makes me unfit to write the history of European collaboration on some area of AI research as originally intended for this collection. However, by interpreting the topic rather loosely, I can regard some European philosophers who were interested in Philosophy of mathematics as early AI researchers from whom I learnt much, such as Kant and Frege. Hume's work is also relevant. Moreover, more recent work by Annette Karmiloff-Smith, begun in Geneva with Piaget then developed independently, helps to identify important challenges for AI (and theoretical neuroscience), that also connect with philosophy of mathematics and the future, rather than the history, of robotics. So this paper presents an idiosyncratic survey of a subset of AI stretching back in time, and deep into other disciplines, including philosophy, psychology and biology, and possibly also deep into the future.

Kant, Frege, Mathematics, Representational-Redescription, Empirical knowledge, Necessary truths, Geometry, Robots

1 Kant, Philosophy of Mathematics and AI

Has research in AI contributed to our understanding of mathematical learning, discovery and reasoning? I shall try to show that this question has different interpretations, leading to different answers. At first sight the question may seem pointless, given all the progress in automated theorem proving and the variety of computer-based mathematical tools used all round the world every day.[1] Yet, as far as I know, there is not yet an AI system that comes close to replicating the kind of learning about the environment that develops into mathematical learning in young children and must in the distant past have happened without the aid of teachers, initially long before the currently used mathematical formalisms were available (Sloman, 2010).

We need new models to help us understand how mathematical competences evolved and developed, and to help us answer philosophical questions about how mathematical knowledge is related to and differs from other kinds. That's the problem that led me into AI over 40 years ago. I'll try, in what follows, to explain the nature of that quest and how it relates to some past thinkers and to past and, I hope, future achievements of AI, closely linked to future advances in biology. One consequence may turn out to be far more intelligent robot "babies" learning about the world, while they learn to think.

[1]For a sample of relevant evidence for, see http://en.wikipedia.org/wiki/Automated_theorem_proving, http://en.wikipedia.org/wiki/Automated_reasoning, http://en.wikipedia.org/wiki/Computational_geometry, http://en.wikipedia.org/wiki/Scientific_computing and other papers in this collection in *AI Communications*.

After a degree in mathematics and physics (1956) I intended to continue into mathematical research, but instead acquired a taste for philosophy and started attending lectures and graduate seminars in Oxford, where I found that most philosophers there thought that a belief often attributed to David Hume (mistakenly according to (Atkinson, 1960)) was correct, namely that there are only two kinds of knowledge, the empirical kind and a kind that merely expresses "relations between ideas" (analytic, essentially trivial knowledge, true by definition). I knew, from my own experience of learning and doing mathematics, that that was false: mathematical knowledge is neither empirical[2] nor trivial. A child who first learns that the result of counting is independent of the order of counting, or that a loosely linked planar polygon with fixed length sides will be rigid if there are only three sides, but not with four, or that changing the height of a planar triangle made of a rubber band stretched by three pins alters its area, whereas moving a vertex parallel to the opposite side does not – is learning something deep, acquiring new knowledge, and is not learning something that could turn out false if the triangle boundary is made of a new kind of material, or has a different colour.[3]

I was also convinced that the view of another, more recent, philosopher, Wittgenstein, was also incorrect: namely that instead of making discoveries, mathematicians merely take decisions that remove residual indeterminacy in their concepts[4], a view similar to that of the great mathematician Henri Poincaré. Another view of mathematics widely discussed at that time was based on the work of Frege, Russell and Whitehead, namely the view that all mathematical knowledge was essentially knowledge of *logical* truths, though that did not make them trivial, since logical discoveries could be deep and unobvious. Frege thought this applied to arithmetical knowledge because arithmetical concepts could be defined in purely logical terms (Frege, 1950), but did not apply to geometric knowledge since the concepts of point, line, area, length, etc. could not.

Russell agreed with Frege about arithmetic, and half agreed with him about geometry, claiming instead in (Russell, 1917) (though he held different views at different times) that all we really can know about geometry and many other branches of mathematics is that IF the axioms are true of anything THEN the theorems are also, because the derivations can all be based on logic. Objections to that view included (a) the possibility of knowing that the Euclidean axioms (possibly with the exception of the parallel axiom and its consequences) are true without deriving them from other axioms, and (b) the widespread use of non-logical methods of proof in geometry and other branches of mathematics, e.g. reasoning with diagrams.

I learnt about philosophical theories about mathematics by attending philosophy lectures, and lectures on foundations of mathematics by Michael Dummett and Hao Wang (my supervisor for a while, though he was mostly in the USA working for IBM on automatic theorem proving[5]). But I felt they were all missing something.

On reading Kant (Kant, 1781), I felt that he understood the nature of mathematical knowledge better than the other philosophers I had encountered (alive or dead).[6] . He claimed that mathematical knowledge was both synthetic (non-analytic) and *a priori* (non-empirical), though not infallible, and not innate, since newborn babies lack the concepts required as well as the ability to make mathematical discoveries. Unsurprisingly (in 1781), he had no precise ideas about the transitions required for a neonate to develop mathematical competences, though it could not be merely a matter of discovering reliable correlations: or the type of statistical learning that now, dominates much of AI, which I believe diverts attention from more important mechanisms discussed below.

So I switched to philosophy, and wrote a thesis defending Kant's philosophy of mathematics (Sloman, 1962). The pressure to publish that now ruins the research of young academics did not exist then, so while teaching various aspects of philosophy in the following few years I continued trying to improve my account of

---

[2]Lakatos (Lakatos, 1976) argues mathematics is *quasi empirical*, since mathematicians can make mistakes, detect them, and correct them.

[3]`http://tinyurl.com/CogMisc/triangle-theorem.html`

[4]This interpretation of his philosophy of mathematics is not agreed by all, but that's not important for this paper.

[5]`http://en.wikipedia.org/wiki/Hao_Wang_(academic)`

[6]There seems to be an intuitive geometry not including anything infinite, subsuming much of Euclidean geometry (without the parallel axiom, but allowing shape-preserving translations or rotations) and usable in reasoning.(Merrick, 2006).
See also `http://tinyurl.com/BhamCog/misc/p-geometry.html`

mathematical knowledge, including publishing papers, e.g. (Sloman, 1965, 1968/9.), which left me feeling that I was missing something important.

In 1969 Max Clowes, an AI vision researcher (Sloman, 1984), arrived at Sussex University, after a spell in Australia. He introduced me to programming and AI, and lent me AI books and papers. I tried to persuade him that there were connections between requirements for making machines see and enabling them to make mathematical discoveries in a non-logical manner, e.g. using diagrams to reason with. I began to suspect that it should be possible to build a "baby" robot that explored the environment like a child, making empirical discoveries but also, after acquiring the required meta-cognitive capabilities, finding that there is a pervasive structure to occupants of space and time, which allows some facts to be *worked out* without having to be demonstrated empirically or inferred statistically.

Similar discoveries could be made about counting and logical reasoning, discoveries triggered by experience then later understood non-empirically, e.g. by noticing invariant properties of counting preserved by permuting things counted, like a child noticing that counting fingers from left to right gives the same result as counting right to left then later realising that the results *must* be the same.[7] That now seems to me to be an example of what Karmiloff-Smith called "Representational Redescription" in (Karmiloff-Smith, 1992).

Many psychologists studying number competences fail to distinguish *numerosity* (roughly: a product of density and area or volume) and *cardinality* (defined in terms of one-to-one correspondence). A robot would require different mechanisms for these. Differences in numerosity become harder to detect as numerosity increases, whereas detection is irrelevant to cardinality: it requires a mapping operation. Numerosity is a very much shallower concept. Most also fail to distinguish understanding metrical concepts from understanding networks of partial orderings (of length, angle, speed, area, volume, force, weight, etc.): I suspect only the latter are available to young children and most other animals.

If we can design a "child robot" that, as a result of interacting with a rich environment while deploying innate metacognitive tendencies (McCarthy, 2008; Chappell & Sloman, 2007)), develops concepts and techniques for perceiving, using and reasoning about affordances (Gibson, 1979; Sloman, 1996a), and is thereby enabled later on to discover necessary, non-empirical, truths about spatial structures, processes and affordances, that would provide a deeper and richer basis for Kant's philosophy of mathematics than anything philosophers have offered. The workings of such a child robot mathematician would be understood if we had designed it. Some of these ideas were later presented in my 1978 book. But that was preceded by momentous events for me, starting in 1971.

2 IJCAI-2, London 1971

Max Clowes was one of the organisers of the 2nd IJCAI in London.[8]. He pressed me to write a paper on my criticisms of the logicist approach to AI, presented by McCarthy and Hayes in 1969 (McCarthy & Hayes, 1969). This also hinted at doubts about the claim that it is possible to assemble *all* the operations used in mathematical reasoning from operations of a Turing machine.

Turing machines can, surprisingly, produce sequences of transformations of symbols that accurately model (at a certain level of abstraction) a wide range of human thought process involving numbers, algebra, and logical formulae, but it is not obvious that those processes suffice for all human intelligence, especially certain kinds of mathematical reasoning about our environments.

Humans can perceive and think about continuously moving geometrical structures changing location, orientation or shape (Sloman, 1982). They can also think about *classes* of such structures and processes without having to represent any *instance* in full detail – unlike physics engines used in computer games, which can generate detailed simulations of particular physical processes, but cannot reason about *classes* of cases as

[7]Chapter 8 of (Sloman, 1971) began to develop that idea `http://tinyurl.com/BhamCog/crp/chap8.html`

[8]http://ijcai.org/Past%20Proceedings/IJCAI-1971/CONTENT/content.htm

humans can, e.g. arguing that a particular operation with ruler and compasses will bisect *any* angle.

The IJCAI paper (also published that year in AIJ) drew attention to modes of representation and reasoning that appeared to fit neither logical and algebraic derivations nor the kinds of atomic operations found in Turing machines. It led Bernard Meltzer to use research council funds to invite me to spend 1972-3 in Edinburgh in his Department.

There I learnt to program in POP-2, was taught Lisp by Boyer and Moore, and got to know outstanding local and visiting AI researchers and PhD students, who kindly gave me some of their time, including several visiting AI researchers. Pat Hayes, then at Essex university, wrote a rejoinder to my IJCAI paper (Hayes, 1984). People I learnt from included Bruce Anderson, Harry Barrow, Danny Bobrow, Bob Boyer, Chris Brown, Frank Brown, Alan Bundy, Julian Davies, Geoffrey Hinton, Bob Kowalski, Christopher Longuet-Higgins, Bernard Meltzer, J Moore, Seymour Papert, Robin Popplestone, Steve Salter, Jim Stansfield, Austin Tate, Sylvia Weir, and others, but I was not able to persuade anyone to join me in trying work out how to build a baby robot that could learn about space and numbers and recapitulate some of the human processes of mathematical discovery, though for a short time Alan Bundy explored the use of diagrammatic reasoning in arithmetic (Bundy, 1973), a topic later pursued by some of his students (mentioned below).

One of the things I learnt that year, especially from limitations of the AI systems I encountered, was the importance of a *complete* architecture combining many kinds of functionality - as proposed in a technical report later included as chapter 6 of (Sloman, 1978). That required all the components to "scale out", i.e. be capable of being combined with other sub-systems to provide new more complex functions. ("Scale out" was a label I used much later, e.g. (Sloman, 2006), to contrast with "scale up".) At that time (early 1970s) most AI research seemed to me to ignore the problems of how to put sub-systems together in a complete architecture embedded in an environment with which it has many concurrent forms of interaction, while also engaged in processing information unrelated to the current environment (a requirement that has been mis-described by proponents of embodied cognition, enactivism and the so-called extended mind who all seem to focus on a subset of what needs to be explained). Of course, computational machinery available at that time had dreadful limits: kilobytes of memory, CPU clock-speeds measured in kiloherz, etc., severely constraining what was possible.

Architectures have received much attention since then, but most researchers merely propose a particular architecture, instead of formulating a general theory allowing a large variety of different architectures (from microbes upwards) to be compared and contrasted – as required for AI to contribute to a scientific or philosophical understanding of biological evolution, and also to avoid short-sighted engineering decisions.[9]

Not much progress was made on ways of representing spatial information so as to model human (animal?) reasoning, though powerful techniques for using logic and other mathematical formalisms were being developed. An example is being able to answer the question: if a plane contains a circle and a triangle, each stretchable and movable, how many points of intersection of the boundaries can there be? Obvious answers are 0 and 6, but most people are able to provide all the other possibilities. If asked how, most talk about visualising various configurations, though one person claimed to start with a set of axioms and derive a theorem! Most people seem to be able to use diagrams (on paper, or merely imagined) to explore both a set of *possibilities* and also *necessary* features of (constraints on) those possibilities, and without requiring precise measurements, nor an infinite set of diagrams to cover infinitely many possibilities.

The combination of requirements (covering possibilities and establishing necessities) is not met by systems that merely produce an example drawing and read off its properties, which in some cases may give the appearance of establishing something if the diagram is interpreted in the right way (e.g. ignoring lengths and angles), but that will not work for the circle and triangle problem, as there are qualitatively distinct subclasses of possibilities to be covered. [10] The work of Lakatos (Lakatos, 1976) suggests that no method can be *guaranteed* to be valid.

Since then, attempts have been made, in Europe and elsewhere, to model various forms of diagrammatic

[9]See the Meta-Morphogenesis project: `http://tinyurl.com/M-M-Gen`

[10]More examples involving triangles are presented in `http://tinyurl.com/CogMisc/triangle-theorem.html`

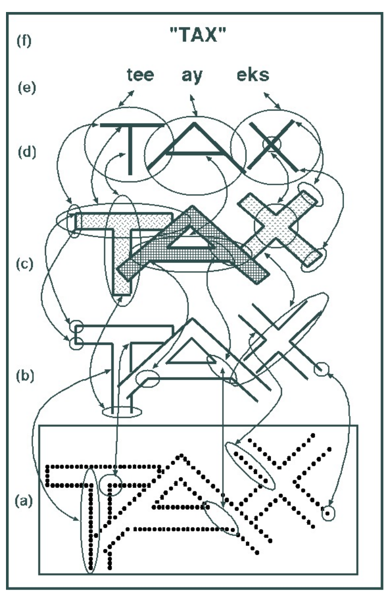
reasoning and a series of international conferences on the topic started in 1998, with results presented in books and journals, e.g. (Anderson, Meyer, & Olivier, 2001). I contributed a paper to the first one, but it merely presented examples of things humans could do, and made no progress on implementation issues. Around that time one of Alan Bundy's students Mateja Jamnik, completed a thesis on diagrammatic reasoning (Jamnik, Bundy, & Green, 1999) (provoked by a lecture in which Roger Penrose claimed that machines could not reason spatially as human mathematicians do!).

While interesting and important, that work, like much else in the field, is concerned only with the use of discrete spatial structures such as 2-D grids, trees or lattices. An exception was Daniel Winterstein's thesis (Winterstein, 2005), though he was more concerned with tutorial support for mathematics students than providing a spatial reasoning system for robots. Perhaps his ideas could be developed in that direction, but I don't think anyone has taken this up. Another strand of research, exemplified by Forbus' work on CogSketch, provides tools to support human diagrammatic reasoning but does not (as far as I know) do the sort of reasoning described here. Likewise Tarski's World[11] uses a diagrammatic display to help a student test logical inferences, but the system does not do any geometric reasoning. (Piaget thought about similar problems (Piaget, 1981, 1983), though not in computational terms.)

There has been much work (e.g. in the Edinburgh AI group) on automated reasoning about discrete computational structures and processes, a simple example of which concerns a recursive procedure (called "*rotate*") which given a list produces another by removing the first item in the list and adding it at the end, so that *rotate([a b c])* is *[b c a]*. It is obvious to humans that applying rotate to L, the length of L times, recreates the original list, L.

A typical AI proof of that would use mathematical induction, whereas (some) humans *visualise* the process of moving the first element of a list to the end once for each element of the list and find it obvious that that *must* create a list with the original elements in the original order. The human ability to use an abstract *spatial* structure (on paper or imagined) to reason reliably about a huge variety of possible instances of a theorem (e.g. all possible list-lengths and all possible contents for each length) has not yet, as far as I know, been modelled, although modes of proof using mathematical induction have.

3 Research on vision



[11] http://ggweb.stanford.edu/tarskisworld

Not long after returning to Sussex I obtained a research council grant to work on vision, developing ideas about architectural requirements for a system that could process visual information at different levels of abstraction in parallel. David Owen, then later Geoffrey Hinton, worked on the project, and Frank O'Gorman, an outstanding programmer who worked with Max Clowes helped. The main features of the system we developed (POPEYE) are described in (Sloman, Owen, Hinton, & O'Gorman, 1978) and Ch 9 of (Sloman, 1978)[12], and illustrated by the figure above, indicating the variety of ontologies used in parallel (bottom up, top-down and middle-out) while interpreting a dot-array ((a) in the diagram).

A paper on POPEYE proved unpublishable, perhaps because David Marr was thought to have demonstrated (a) that sophisticated architectures are not required for human vision, (b) that problems of interpretation using artificial images do not arise for natural scenes because the richness of natural scenes rules out the ambiguities of artificial images, such as line drawings. For similar reasons (I think), an application to continue the project was rejected. So theoretical work continued informally and unfunded for several years, leading to papers on vision, architectures and representation, including the need for intelligent systems to use multiple forms of representation. (See (Sloman, 1982, 1989, 1993, 1995, 1996a, 1996b, 2000, 2002, 2001, 2005, 2007).) The investigation of architectures led to work on emotional and other affective states and processes and to design of test domains to challenge current theories of vision and spatial reasoning.

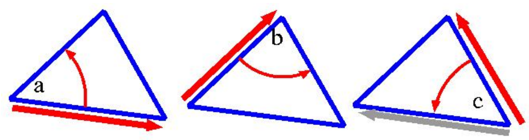
The key challenges were concerned with perception of and reasoning about continuous changes producing new structures, use of continuously changing visual information for controlling action, perception of empty spaces (especially potentially useful empty spaces), perception of unrealised possibilities for spatial changes (whether actions of the perceiver or not – contra Gibson), and perception of constraints on possibilities. None of these is captured by graphical simulations, or game engines. Instead, an intelligent perceiver requires a level of abstraction that corresponds to mathematical reasoning about possibilities and constraints (e.g. a construction for bisecting *any* angle). As far as I know the problems are still unsolved, though there are fragmentary steps in the required direction, e.g. (Cabalar & Santos, 2011; Li & Cohn, 2012; Forbus, Usher, Lovett, Lockwood, & Wetzel, 2011)

The ontology of POPEYE did not include features required for mathematical meta-cognition, namely representations of possible changes ("proto-affordances") and their implications/invariants, required to understand the interactions between worm and pinion in this sort of configuration (from Wikimedia):

For example, why does rotation of the worm about its axle make the pinion rotate about its axle? What about the reverse influence?

In this mode of reasoning, diagrams and transformations of diagrams play a different role from the data-structures and variables representing spatial relationships and processes in physical simulation engines required for weather forecasting and game engines. Those represent geometrical structures and processes at a high level of precision and allow only predictions at that level of precision, possibly weakened by probability distributions, whereas a mathematician typically uses a diagram, and geometric or topological alterations to the diagram, to represent a whole class of structures and processes, and their invariants. Here is a proof (due to Mary Pardoe) that the angles of a planar triangle add up to half a rotation (180 degrees).

[12] http://tinyurl.com/crp78/chap9.html



The proof is a process, in which the red arrow rotates through angles a, b, and c, then ends up on the original line pointing in the opposite direction. This, like many proofs in Euclidean geometry, does not depend on the specific sizes of the angles or the lengths of the sides. Rather it reveals a collection of "invariant" relationships common to a large class of cases (all shapes, sizes and orientations, of triangles, provided that they are planar). This is quite different from the use of a physics engine to predict the outcome of a skid on a race-track, or a building collapsing (whether the prediction is deterministic or probabilistic).

A related lack of generalisation would be manifested by a humanoid robot trained to use one hand to rotate a crank handle at a fixed rate, which has to be re-trained in order to perform the action after its body is moved a short distance in relation to the crank. If, instead of learning a specific set of sensory-motor relationships, it understood that the action required constantly pushing the crank handle in the (constantly changing) direction of the tangent of the circle through which it moves, then it might be able to constantly work out the direction of the force it needs to apply to the handle, and work backwards from that to the motor signals required. This is different from learning for each configuration of sensors and motors how to produce the next motor signals: that uses a different ontology.

I suggest that the human ability to prove theorems in Euclidean geometry arose out of more general abilities to reason in an abstract way about possibilities and constraints in physical structures and processes.

4 The EU cognitive systems project

The EU Cognitive systems project, announced in the summer of 2003, provided an important new opportunity to bring together researchers in robotics, psychology, philosophy, mathematics, education, and other fields, especially as the emphasis was mainly on science, as opposed to mere engineering.[13] But progress was held up by the pressure to produce working systems to demonstrate to reviewers every year, along with widespread, and in my view unreflective, commitment to a shallow notion of "embodied cognition", criticised in (Sloman, 2009). These pressures and commitments seemed to get in the way of more reflective research on *long term requirements* for intelligent machines which, instead of always being engaged in acting on their environments (using on-line intelligence), can also reason about, explain, form theories about, and form plans for action in, their environments without necessarily doing anything in the environment at the time (using off-line intelligence).

The narrowly focused view of embodiment also impacted on AI research on vision. There was a lot of work on giving machines the ability to classify images and parts of images, and to combine information from images to form depth-maps, and to use vision to control various sorts of motion, e.g. catching a moving object, and (in SLAM systems) to form large scale maps from terrain fragments sampled consecutively, but, as far as I know, there were few, if any, attempts to model increasing understanding of spatial structures and processes in a way that included perception of possibilities and constraints inherent in a class of spatial configurations, especially "multi-strand" relationships and processes described in (Sloman, 2009).

Some researchers confuse that spatial reasoning capability with the ability to run a simulation model in order to predict what will happen, as occurs in game engines – which cannot answer questions like "Why couldn't the block pass through the hole?" by responding "Because it was in the wrong orientation". Nor can they explain why rotating a threaded nut on a fixed threaded bolt causes the nut to move along the length of the bolt, in a direction that depends on the direction of rotation and the screw thread direction, even if they can predict the rotation for every possible nut+bolt combination. A generic explanation need not mention the

[13] `http://cognitivesystems.org/events/CogSysKickoff/files/unitE5.pdf`

specific dimensions involved, e.g. the diameter of the bolt, the depth of the screw thread, the pitch of the thread, the speed of rotation, etc., whereas physical simulation packages cannot run without having precise measures for all relevant physical quantities.

Some AI systems learn about affordances by being *told* by a human, for example, that grasping is afforded by the handle of a cup, but, as far as I know, there is no machine that uses perception of relationships between the shape of a cup with a handle and the structure and possible movements of hands and fingers to infer positive and negative affordances. A machine might learn from a lot of training examples that the handle of a cup can be used to lift the cup, but different mechanisms are required to be able to explain *why* lifting a cup full of coffee by pushing a finger through the handle and lifting the finger is a bad idea, even though it will raise the cup off the table – i.e. explaining positive and negative affordances involved in the same action.

The kind of reflective understanding of possible changes, and constraints on changes, in a physical situation, summarised in (Sloman, 1996a), and required for developing human-like mathematical competences, is not available to any robot I've heard of, though there are many that can perform impressive physical actions, like swimming, hopping, catching a ball, for example, and others that, if given an appropriate collection of factual premises about a situation in a planning formalism, can solve complex planning tasks, including solving spatial/topological puzzles as in (Cabalar & Santos, 2011).

What seems to be missing in AI systems is the ability to discover a collection of useful generalisations empirically which are then replaced by a "generative" theory that allows new cases to be derived from the theory instead of having to be learnt through a training process.

5 Links with developmental neuroscience

I think there is a deep connection between (a) the work of Annette Karmiloff-Smith (K-S), a neuro-developmental psychologist, reported in (Karmiloff-Smith, 1992), especially her ideas about "Representational Redescription" (RR), and (b) the problem of using AI designs to defend Kant's philosophy of mathematics, according to which mathematical knowledge is neither empirical nor trivial (e.g. true by definition). The kind of reorganisation of information K-S discussed (RR) seems to be closely related to the Kantian notion that a child can learn things about the world by playing and experimenting in it, and can then, without necessarily being instructed by a teacher, reorganise that knowledge so that (i) it is no longer empirical but derivable within a generative system (using still unknown mechanisms) and (ii) it can be extrapolated to many new configurations without those configurations having to be tested experimentally.

Such changes may require not only reorganisation of empirically acquired information but also ontology expansion and a modified information-processing *architecture*, enabling use of a generative rule-set. The best known example of RR that K-S discusses is the transformation in children from acquiring what seems to be a *pattern-based* ability to use language to a *syntax-based* ability, which provides much greater generative power. After that change, children start making grammatical mistakes, e.g. saying "He runned away" instead of "He ran away".

Although I don't think anyone knows exactly what goes on in the child's brain it is clear that for a computational system designer a change in the software architecture is required to replace a pattern based system (using fixed patterns) with a (recursive) generative grammar based system. Likewise extending a generative system to cope with counter-examples (e.g. using "ran" as past tense of "run") requires an extension of the architecture. Neither change is merely a change of formalism: new mechanisms are needed.

The mechanisms K-S postulates for reorganising a behavioural competence acquired empirically to form a more generative (or deductive) competence may turn out to be closely related to the biological mechanisms that led to the discovery of truths of Euclidean geometry and their organisation into a generative system in Euclid's *Elements*, though I am no historian of mathematics. RR also seems to be closely related to Craik's ideas in (Craik, 1943) regarding animals using models to work things out instead of trying them out, though I don't know if he noticed the importance of reasoning about *sets of possibilities* as opposed to predicting *specific* *behaviours*. In chapter 8 of (Sloman, 1978)[14] I tried to indicate some of the ways in which counting expertise might be re-organised as a result of playing with counting and building up new re-usable structures. But the ideas were very sketchy and never implemented in a working system.

The label "toddler theorem" is proposed for the type of generic result that young children seem to be able to derive after reorganising knowledge acquired by experimenting with instances. In older children, with a good grasp of cardinality, knowledge acquired by playing with blocks could lead to the discovery that some numbers are prime, some not.[15]

Future research linking ideas of K-S with problems in philosophy of mathematics, in the context of a general theory of how biological evolution constantly produces new forms of competence substantially extending or transforming previous competences, may help to remove the existing gaps in the explanatory power of AI models of mathematical reasoning. This will also have implications for empirical research in biology, neuroscience and education as well as applications in intelligent robots that learn new ways to work things out for themselves.

K-S distinguishes several stages of representational re-description, relevant in different ways to this paper, including a stage in which agents become aware of and able to think about how a reasoning task is performed, a stage in which such meta-knowledge is expressed explicitly in words or thoughts, or communicated to others, and possibly further stages, including being able to assess the type of competence reached by another individual (e.g. a child) and being able to intervene and help that individual make progress. Implementing all this will require significant modifications to current AI research on meta-cognition, e.g. (Cox & Raja, 2011).

Until we learn how to implement mechanisms (including new forms of representation) capable of producing the sorts of transition discussed by K-S and used in informal but substantive mathematical discoveries, AI in general and robotics in particular will continue making promises, setting up challenges, but only making incremental, relatively shallow, though sometimes deceptively impressive, progress, such as the Boston Dynamics BigDog robot.[16]

Although I cannot identify the required mechanisms, I have tried to describe the problem (though there is much more to be said), and in various web pages have presented examples in the form of challenges (e.g. (Sloman, 2008)), with examples to illustrate some of the transitions in understanding that can occur. It should be possible for AI researchers and others to work collaboratively on identifying a much richer collection of requirements for a satisfactory solution, and a larger collection of intermediate stages in evolution between microbes and crows, squirrels, elephants, octopuses, and humans.[17] Of course, there may already be researchers doing what I claim needs to be done, whom I have not encountered.

6 Conclusion

The label 'Meta-morphogenesis' can be used for a broad multi-disciplinary research programme, linking evolution, development, learning, and problems in AI/Robotics, for which ideas are being collected, e.g. in `http://tinyurl.com/M-M-Gen`. See also (Sloman, 2013 In Press). I suspect this will lead to new contributions to philosophy of mathematics, biological investigations of evolution of various kinds of intelligence, psychological investigations of various kinds of learning and development, neuroscientific investigations of brain mechanisms involved in these processes, epigenetic investigations of how a genome is able to encode the right sort of constrained potential for development, e.g. as found in altricial species, research into the role of cultural evolution in extending individual mental capabilities, and perhaps also ways of using some of the results of all that research to design new more intelligent robots and other machines. A conjecture

[14] `http://tinyurl.com/crp78/chap8.html`
[15] `http://tinyurl.com/TodTh#primes`
[16] `http://www.youtube.com/watch?v=cNZPRsrwumQ`
[17] `http://tinyurl.com/CogMisc/evolution-info-transitions.html`

that evolution discovered the importance of self-monitoring virtual machinery long before we did may play an important role in explaining mathematical meta-cognition and other aspects of human consciousness.

NOTE 1: I object to closed-access publication of research results, but this journal allows freely available copies to be posted by authors. This paper is a partial, highly compressed report on a personal journey over the last four decades. Missing details and references will be added in a longer, more complete, version updated from time to time on this web site: `http://tinyurl.com/BhamCog/13.html#ecai` .

NOTE 2: A more recent paper discussing the nature of mathematics and the evolution of mathematical competences is available in draft form at
`http://www.cs.bham.ac.uk/research/projects/cogaff/incomputable-sloman.pdf`
"Evolution of Geometrical Reasoning"

References

Anderson, M., Meyer, B., & Olivier, P. (Eds.). (2001). *Diagrammatic representation and reasoning*. Berlin: Springer-Verlag.

Atkinson, R. F. (1960, Apr). Hume on Mathematics. *Phil. Quart.*, *10*(39), 127–137. Available from http://www.jstor.org/stable/2960061

Bundy, A. (1973). Doing arithmetic with diagrams. In *Proc 3rd ijcai* (pp. 130–138). Morgan Kaufmann Publishers Inc. Available from ijcai.org/Past\%20Proceedings/IJCAI-73/PDF/015.pdf

Cabalar, P., & Santos, P. E. (2011). Formalising the fisherman's folly puzzle. *Artificial Intelligence*, *175*(1), 346–377. Available from http://www.sciencedirect.com/science/article/pii/S0004370210000408 (John McCarthy's Legacy)

Chappell, J., & Sloman, A. (2007). Natural and artificial meta-configured altricial information-processing systems. *International Journal of Unconventional Computing*, *3*(3), 211–239. Available from http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0609

Cox, M. T., & Raja, A. (Eds.). (2011). *Metareasoning: Thinking about thinking*. Cambridge, MA: MIT Press.

Craik, K. (1943). *The nature of explanation*. CUP.

Forbus, K., Usher, J., Lovett, A., Lockwood, K., & Wetzel, J. (2011). Cogsketch: Sketch understanding for cognitive science research and for education. *Topics in Cognitive Science*, *3*(4), 648–666.

Frege, G. (1950). *The Foundations of Arithmetic: a logico-mathematical enquiry into the concept of number*. Oxford: B.H. Blackwell. ((Tr. J.L. Austin. Original 1884))

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

Hayes, P. (1984). Some problems and non-problems in representation theory. In R. Brachman & H. Levesque (Eds.), *Readings in knowledge representation* (pp. 3–22). Los Altos, California: Morgan Kaufmann.

Jamnik, M., Bundy, A., & Green, I. (1999). On automating diagrammatic proofs of arithmetic arguments. *Journal of Logic, Language and Information*, *8*(3), 297–321.

Kant, I. (1781). *Critique of pure reason*. London: Macmillan. (Translated (1929) by Norman Kemp Smith)

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: MIT Press.

Lakatos, I. (1976). *Proofs and Refutations*. Cambridge: CUP.

Li, S., & Cohn, A. G. (2012). Reasoning with topological and directional spatial information. *Computational Intelligence*, *28*(4), 579–616.

McCarthy, J. (2008). The well-designed child. *Artificial Intelligence*, *172*(18), 2003–2014. Available from http://www-formal.stanford.edu/jmc/child.html

McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of AI. In B. Meltzer & D. Michie (Eds.), *Machine Intelligence 4* (pp. 463–502). Edinburgh, Scotland: Edinburgh University Press. Available from http://www-formal.stanford.edu/jmc/mcchay69/mcchay69.html

Merrick, T. (2006). What Frege Meant When He Said: Kant is Right about Geometry. *Philosophia Mathematica*, *14*(1), 44–75. (doi:10.1093/philmat/nkj013)

Piaget, J. (1981). *Possibility and Necessity Vol 1. The role of possibility in cognitive development*. Minneapolis: U. of Minnesota Press. (Tr. by Helga Feider from French in 1987)

Piaget, J. (1983). *Possibility and Necessity Vol 2. The role of necessity in cognitive development*. Minneapolis: U. of Minnesota Press. (Tr. by Helga Feider from French in 1987)

Russell, B. (1917). *Mysticism and logic and other essays*. London: Allen & Unwin.

Sloman, A. (1962). *Knowing and Understanding: Relations between meaning and truth, meaning and necessary truth, meaning and synthetic necessary truth*. Unpublished doctoral dissertation, Oxford. Available from `http://tinyurl.com/BhamCog/07.html#706`

Sloman, A. (1965). 'Necessary', 'A Priori' and 'Analytic'. *Analysis*, *26*(1), 12–16. Available from `http://tinyurl.com/BhamCog/07.html#701`

Sloman, A. (1968/9.). Explaining Logical Necessity. *Proceedings of the Aristotelian Society*, *69*, 33–50. Available from `http://tinyurl.com/BhamCog/07.html#712`

Sloman, A. (1971). Interactions between philosophy and AI: The role of intuition and non-logical reasoning in intelligence. In *Proc 2nd ijcai* (pp. 209–226). London: Kaufmann. Available from `http://tinyurl.com/BhamCog/04.html#200407`

Sloman, A. (1978). *The computer revolution in philosophy*. Hassocks, Sussex: Harvester Press. Available from `http://tinyurl.com/BhamCog/crp`

Sloman, A. (1982). Image interpretation: The way ahead? In O. Braddick & A. Sleigh. (Eds.), *Physical and Biological Processing of Images (The Rank Prize Funds Conference, 1982)* (pp. 380–401). Berlin: Springer-Verlag. Available from `http://tinyurl.com/BhamCog/06.html#0604`

Sloman, A. (1984). Experiencing Computation: A Tribute to Max Clowes. In M. Yazdani (Ed.), *New horizons in educational computing* (pp. 207–219). Chichester: Ellis Horwood. Available from `http://tinyurl.com/BhamCog/06.html#0604`

Sloman, A. (1989). On designing a visual system (towards a gibsonian computational model of vision). *Journal of Experimental and Theoretical AI*, *1*(4), 289–337. Available from `http://tinyurl.com/BhamCog/81-95.html#7`

Sloman, A. (1993). The mind as a control system. In C. Hookway & D. Peterson (Eds.), *Philosophy and the cognitive sciences* (pp. 69–110). Cambridge, UK: CUP. Available from `http://tinyurl.com/BhamCog/81-95.html#18`

Sloman, A. (1995). Musings on the roles of logical and non-logical representations in intelligence. In J. Glasgow, H. Narayanan, & B. Chandrasekaran (Eds.), *Diagrammatic reasoning: Computational and cognitive perspectives* (pp. 7–33). Cambridge, MA: MIT Press.

Sloman, A. (1996a). Actual possibilities. In L. Aiello & S. Shapiro (Eds.), *Principles of knowledge representation and reasoning: Proc. 5th int. conf. (KR '96)* (pp. 627–638). Boston, MA: Morgan Kaufmann Publishers. Available from `http://tinyurl.com/BhamCog/96-99.html\#15`

Sloman, A. (1996b). Towards a general theory of representations. In D.M.Peterson (Ed.), *Forms of representation: an interdisciplinary theme for cognitive science* (pp. 118–140). Exeter, U.K.: Intellect Books.

Sloman, A. (2000). Models of models of mind. In M. Lee (Ed.), *Proceedings of Symposium on How to Design a Functioning Mind, AISB'00* (pp. 1–9). Birmingham: AISB. Available from `http://tinyurl.com/BhamCog/00-02.html#56`

Sloman, A. (2001). Evolvable biologically plausible visual architectures. In T. Cootes & C. Taylor (Eds.), *Proceedings of British Machine Vision Conference* (pp. 313–322). Manchester: BMVA.

Sloman, A. (2002). Diagrams in the mind. In M. Anderson, B. Meyer, & P. Olivier (Eds.), *Diagrammatic representation and reasoning* (pp. 7–28). Berlin: Springer-Verlag. Available from `http://tinyurl.com/BhamCog/00-02.html#58`

Sloman, A. (2005). *Perception of structure: Anyone Interested?* (Research Note No. COSY-PR-0507). Birmingham, UK. Available from
http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0507

Sloman, A. (2006, May). *Requirements for a Fully Deliberative Architecture (Or component of an architecture)* (Research Note No. COSY-DP-0604). Birmingham, UK: School of Computer Science, University of Birmingham. Available from
http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0604

Sloman, A. (2007, Nov). *Predicting Affordance Changes (Alternatives ways to deal with uncertainty)* (Tech. Rep. No. COSY-DP-0702). UK: School of Computer Science, Birmingham. Available from
http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0702 (Unpublished discussion paper (HTML))

Sloman, A. (2008). *A Multi-picture Challenge for Theories of Vision* (Research Note No. COSY-PR-0801). (http://tinyurl.com/BhamCog/talks/#talk88)

Sloman, A. (2009). Some Requirements for Human-like Robots: Why the recent over-emphasis on embodiment has held up progress. In B. Sendhoff, E. Koerner, O. Sporns, H. Ritter, & K. Doya (Eds.), *Creating Brain-like Intelligence* (pp. 248–277). Berlin: Springer-Verlag. Available from
http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0804

Sloman, A. (2010). If Learning Maths Requires a Teacher, Where did the First Teachers Come From? In A. Pease, M. Guhe, & A. Smaill (Eds.), *Proc. Int. Symp. on Mathematical Practice and Cognition, AISB 2010 Convention* (pp. 30–39). Available from http://tinyurl.com/BhamCog/10.html#1001

Sloman, A. (2013 In Press). Paper 4: Virtual machinery and evolution of mind (part 3) meta-morphogenesis: Evolution of information-processing machinery. In S. B. Cooper & J. van Leeuwen (Eds.), *Alan Turing - His Work and Impact.* Amsterdam: Elsevier. Available from
http://tinyurl.com/BhamCog/11.html#1106d

Sloman, A., Owen, D., Hinton, G., & O'Gorman, F. (1978, 18-20th). Representation and Control in Vision. In D. Sleeman (Ed.), *Proc. AISB/GI Conference* (pp. 309–315). Hamburg. Available from
http://tinyurl.com/BhamCog/62-79.html#1978-03

Winterstein, D. (2005). *Using Diagrammatic Reasoning for Theorem Proving in a Continuous Domain.* Unpublished doctoral dissertation, University of Edinburgh, School of Informatics. Available from
http://www.era.lib.ed.ac.uk/handle/1842/642